Ch 3, Data Analytics: Drawing Conclusions, Part 2

# Identifying patterns and relationships between data, 152-165

1. **What is the purpose of using statistics when working with large data sets?**

Statistics convert large quantities of raw data into smaller, informative and meaningful summaries, allowing patterns and relationships to be easily identified.

1. **List the 3 most useful statistical concepts from table 3.7, p 152.**

|  |  |
| --- | --- |
| Average | The single value that gives the most representative summary value of a range of numbers; mean, mode and median are different ways of averaging data under different circumstances. |
| Standard Deviation | Indicates how consistent the values are in a data set. Are the values all close to the average, or do they vary greatly? A low standard deviation means that the mean is truly representative of all the data and can be trusted. A high standard deviation suggests you should not trust the mean. |
| Correlation | Do two data sets show the same trends, as if they might be connected in some way? |

# Basic statistics

1. **Distinguish between the three methods of calculating an average.**

The methods of calculating an average include the mean, median, and mode.

1. **What is the purpose of the standard deviation?**

Standard deviation measures how much a data set varies from the mean of the data.

# Correlation and causality, p 156

1. **Why do you need to be careful when looking at patterns in data in terms of cause and effect?**

You need to be careful when looking at patterns as the human brain is hard-wired to look for patterns in information, this tendency can be misleading as we may visualize a pattern that may not exist.

# Data visualisations, p 157

1. **What are data visualizations used for?**

Data visualizations is the representation of (digital) numbers through use of (analogue) lines, shapes and colors making it easier to interpret the data.

# Queries and searches, p 159

1. **How can queries, searches, filtering and sorting be used when analyzing large data sets?**

Queries, searches, filtering and sorting can be used when analyzing large data sets as it handles bulk data by hiding most of the data, highlighting the interesting parts which remain.

# Conditional formatting, p 164

1. **Explain the process of conditional formatting.**

Conditional formatting is changing the appearance of data automatically based on its current value in spreadsheets/databases, highlighting data that may be of special interest, also possibly highlight errors inside the database.

(p164 example process).